In this sample no significant differences in the attitudes to examination were detectable between Social Classes I, II, and VII, and Social Classes V and VI.

(3) ATTITUDES TO HOME VISITS (tables IX-XII)

Patients were asked: "Would you prefer a student *not* to be present when the doctor calls?"

Here women qualified their answers much more frequently than men—often expressing the view that there might be some occasions on which they preferred the doctor to visit unaccompanied. The view, however, did not appear to be affected by previous contact with students, nor by age group or social class (tables IX-XII).

(4) ATTITUDES TO THE PRESENCE OF TWO STUDENTS

Thirty nine patients (15%) expressed the view that they would feel more upset by the presence of two students rather than one —either at consultation or at home visit. This view was more common among women than among men (table XIII). This difference is significant at the 5% level ($\chi^2=6{\cdot}52$; $P<{\cdot}05$).

Implications

These findings suggest that the presence of undergraduate students in general practice carries important implications. Firstly, the presence of a student may complicate the task of eliciting relevant psychosocial components at consultation. This appears to be particularly so with women. Thus a general practitioner must be alert to this possibility when a student is present. Secondly, the general-practitioner teacher must constantly monitor the effects that the student's presence is having on his patient—which may be indicated by much non-verbal communication. Certainly, patients will rarely (if ever) directly ask the doctor if they may consult him alone. Thus in this study 185 patients had previously consulted their doctor when a student was present (84 of them on three or more occasions): yet only three (one man, two women) confessed to ever having asked the doctor to see him alone. When a student is introduced, therefore, perhaps the simplest and most effective practice is for the doctor to ask his patients routinely whether they wish to consult him alone. To adopt this routine goes some way towards avoiding the barriers to communication which may otherwise arise.

Thirdly, given this approach, the general-practitioner teacher may well wish to consider having two students present. Educationally, such an arrangement facilitates learning by the interaction it offers between the students themselves. Nevertheless, patients—without exception—responded unfavourably when questioned about their reactions to the presence of more than two students.

This study was generously supported by a grant from the Research Foundation Board of the Royal College of General Practitioners.

Copies of the interview proforma may be had on request from H.J.W.

References

1 Harris, C. M., (1969). *General Practice Teaching of Undergraduates in British Medical Schools.* Reports from General Practice No. XI. London, Royal College of General Practitioners.
2 Byrne, P. S., *Journal of the Royal College of General Practitioners*, 1973, 23, Suppl. No. 1.
3 Wright, H. J., *Journal of the Royal College of General Practitioners*, 1973. In press.
4 Freeling, P., and Browne, K., (1966), *The Doctor Patient Relationship.* Edinburgh and London, Livingstone.

Computers in Medicine

Human and Computer-aided Diagnosis of Abdominal Pain: Further Report with Emphasis on Performance of Clinicians

F. T. DE DOMBAL, D. J. LEAPER, JANE C. HORROCKS, JOHN R. STANILAND, A. P. McCANN

British Medical Journal, 1974, 1, 376–380

Summary

This paper reports a controlled trial of human and computer-aided diagnosis in a series of 552 patients with acute abdominal pain. The overall diagnostic accuracy of the computer-aided system was 91·5% and that of the senior clinician to see each case was 81·2%. However, the clinician's diagnostic performance improved markedly during the period of the trial. The proportion of appendices which perforated before operation fell from 36% to 4% during the trial, and the negative laparotomy rate dropped sharply. After the trial closed in August 1972 these figures reverted towards their pretrial levels.

It is suggested that while computer-aided diagnosis is a valuable direct adjunct to the clinician dealing with the "acute abdomen," he may also benefit in the short-term from the constant feedback he receives and from the disciplines and constraints involved in communicating with the computer.

University Department of Surgery, General Infirmary, Leeds LS1 3EX

F. T. DE DOMBAL, M.D., F.R.C.S., Reader in Clinical Information Science
D. J. LEAPER, M.B., CH.B., Research Assistant (Now Surgical Registrar)
JANE C. HORROCKS, Computer Programmer and Physician's Assistant
JOHN R. STANILAND, Research Assistant (Present address: Trinity Hall, Cambridge)

Centre for Computer Studies, The University, Leeds LS2 9JT

A. P. McCANN, M.SC., PH.D., Lecturer in Computational Science

Introduction

In recent years the concept of computer-aided diagnosis has been the subject of much research and not a little con-

troversy. In 1971 we instituted such a system, and in studies recently reported,[1 2] discussed its use in a controlled trial involving some 304 patients with acute abdominal pain.

This further report deals mainly with two additional aspects of the study. Firstly, the report outlines the additional experience gained between December 1971 and August 1972 —so that in all the report now analyses a prospective consecutive series of some 552 patients with abdominal pain. Secondly, and more important, this further report attempts to assess the reactions of the clinicians to the system, and investigates the clinical performance of the relevant team during the past three years.

Conduct of Investigation

This investigation comprised a prospective unselected controlled real-time real-life trial involving all patients admitted to the wards of the professorial surgical unit in the Leeds General Infirmary with acute abdominal pain between 1 January 1971 and 31 July 1972. Criteria for admission to the series were somewhat strict, being (1) that the patient's presenting complaint was abdominal pain, (2) that this pain had arisen within one week of admission, (3) that the admission was an emergency procedure via the receiving room, (4) that the patient was not already on the series (ruling out by definition any repeated admissions), (5) that the patient was capable of giving a medical history (ruling out an unaccompanied baby aged 2 weeks, and one further unconscious patient), and (6) that a diagnosis was eventually made (ruling out two or three further patients with equivocal pathological reports after appendicectomy). Even allowing for these strict criteria, however, a total of 552 patients were admitted into the survey.

Much of the protocol of the study has already been reported.[2] We studied the patients alongside the clinical team, noting the diagnosis of each of the clinicians who saw the patient. Details of the case history were fed into the computing system (fig. 1), these data being obtained from the registrar when he first saw each case, and a "real-time" diagnosis was produced before the patient was taken to the operating theatre (if this course of management was deemed appropriate by the clinical team). The computer's predictions and the final (usually operative) diagnosis were noted and filed for later analysis.

Three additional points are worthy of comment. Firstly, we not only paid attention to the operative diagnosis, but also paid considerable attention to the operative findings; these reflected to some extent the efficacy of management of the patient, a point which will be dealt with in detail later. Secondly, we did not at once inform members of the clinical team of the computer's predictions, but we did discuss the case with them once this initial decision and management had been made and undertaken. Certainly after each "take-in" day we spent some time discussing the relevant cases with members of the clinical team, providing them with a delayed form of "feedback" and analysis of their own and the computer-aided system's performance. Finally, aware of the cogent comments of Shepherd,[3] we looked again at the problem of "non-specific abdominal pain"—patients who settled down and went home.

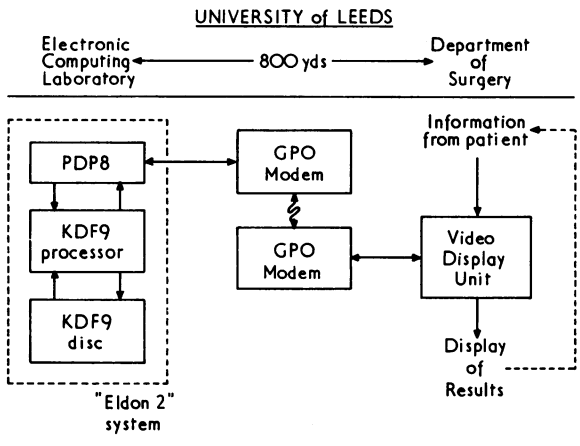

FIG. 1—Block diagram of computer-aided system for clinical diagnosis used in present study. (800 yd = 731 m.)

Results

OVERALL FINAL DIAGNOSES

The final diagnosis made in each of the 552 patients is shown in table I. Most of the categories are self-explanatory, but it may be helpful to provide additional data about some of them, particularly non-specific abdominal pain (N.S.A.P.).

TABLE I—*Final Diagnosis in a Consecutive Prospective Unselected Series of 552 Cases with Acute Abdominal Pain presenting between January 1971 and July 1972*

Disease Category	No. of Cases	Percentage of Total
N.S.A.P.	279	50·5
Appendicitis	145	26·3
Cholecystitis	42	7·6
Small bowel obstruction	20	3·6
Perforated duodenal ulcer	17	3·1
Pancreatitis	16	2·9
Diverticular disease	11	2·0
Other (miscellaneous)	22	4·0

This (N.S.A.P.) as Shepherd aptly remarks[3] is scarcely a diagnosis, and yet it is a concept with which every surgeon faced with an "acute abdomen" is familiar. By N.S.A.P. is meant a condition which falls into one of three categories. Either (1) the patient's pain settles down and the patient goes home with no diagnosis made at all, or (2) the patient undergoes a laparotomy (which shows no disorder), or (3) a diagnosis is made—such as a diagnosis of "urinary tract infection" based on two or three pus cells in a sample of urine —which may or may not be responsible for the pain, but which needs no surgical intervention. While not particularly enchanted with the term N.S.A.P. to cover such contingencies we are unable to think of anything more suitable. N.S.A.P. does, moreover, have the additional merit of embracing a group of patients whose short-term management is similar and non-operative; and finally it should be reiterated that when most were followed up in the outpatients some weeks after their admission we failed to find a single instance of a patient with N.S.A.P. presenting—either here or elsewhere—with proved acute appendicitis in this (admittedly short) period of follow-up.

One other comment about categorization concerns the category "small bowel obstruction." We have included in this category two cases whose obstruction was due to neoplasm of the caecum, since the effective obstruction was in the small bowel, even though the cause of the obstruction was distal to the ileocaecal valve.

Finally, as regards the "other" miscellaneous category, Shepherd[3] commented that in our earlier series the proportion of patients in this category (2·6%) was surprisingly low. Indeed in the present larger series this figure has risen—to around 4%—but the rise has been nowhere near that predicted (to 15%), probably because Shepherd was commenting on the basis of 1,179 *operations* whereas we ourselves report experience with 552 *patients*.

OVERALL FINAL RESULTS

The survey ended in August 1972 at which time some 552 patients had been accepted for analysis. The diagnostic accuracy of various groups of clinicians and the computer-aided system is shown in table II. These in broad measure confirm the trends in our earlier report. It is perhaps interesting, however, to note that the original computer-aided system's diagnostic accuracy level has remained almost exactly the same in this extended series—indicating that the cases seen have varied but little in respect of "difficulty" during the time period covered by the trial.

TABLE II—*Diagnostic Performance of Clinicians and a Computer-aided System during Two Study Periods*

	Diagnostic Accuracy (%)	
	Cases 1-304 (Jan.-Dec. 1971)	Cases 305-552 (Dec. 1971-Aug. 1972)
Diagnosis on admission	44·8	38·9
House surgeons	72·2	69·6
Registrars	77·0	82·1
Senior clinician	79·6	83·1
Computer-aided system	91·8	91·2

COMPARISON OF TWO STUDY PERIODS

Of considerable interest are the data in table II which compares results from two periods, the original one of 11 months up to December 1971 and the subsequent period of eight months up to August 1972. Thus, for example, the original diagnostic accuracy rate of 44·8% on admission actually went *down* to under 39% during the last eight months of our study, as did the accuracy of the house surgeons (by a slight margin, from 72·2% to 69·6%). However, in contrast the overall diagnostic accuracy of the registrars rose from 77·0% in the first half of our trial to 82·1% in the second half, and there was a commensurate rise in the accuracy of the senior clinician to see each case (79·6% to 83·1%).

PERFORMANCE OF COMPUTER-AIDED SYSTEM

There is little to add concerning the performance of the computer-aided system to the remarks in our earlier report, though for interest details of the computer-aided system's performance are given in table III. It may, however, be relevant to look at some of the reasons for computer error in the table.

By far the most common reason for error was the failure of the computing system to deal with cases outside its "remit." That is to say patients with gangrenous ovarian cysts were mis-diagnosed as "appendicitis" and a few further patients with perforated carcinoma of sigmoid were erroneously classified as "acute diverticular disease." We recognize this as a major problem of the system, have designed further programmes to cover both these clinical areas (gynaecology and lower gastrointestinal disorders), and will be reporting on these in due course. The remaining area of computer error (prediction of "appendicitis" in eight cases who underwent negative laparotomy) we are less inclined to tamper with since in discriminating between appendicitis and N.S.A.P. the computer-aided system still has a handy margin of performance over the unaided clinician (fig. 2).

TABLE III—*Outline of Computer-aided System's Performance in Series of 552 Cases*

Disease	No. of Cases	Correctly Diagnosed	Accuracy (%)
Appendicitis	145	141*	97·2
Diverticular disease	11	11	100·0
Perforated duodenal ulcer	17	17	100·0
N.S.A.P.	279	262*	93·9
Cholecystitis	42	42	100·0
Small bowel obstruction	20	19	95·0
Pancreatitis	16	13	81·3
Other (miscellaneous)	22	Nil	Nil
Total	552	505	91·5%

* Four cases (two appendicitis, two N.S.A.P.) listed as "fail-safe", since clinicians could not agree on case history. Computer failed by making no diagnosis at all. Listed and counted as failures. Other two cases of appendicitis diagnosed as cholecystitis (one case) and diverticular disease (one case).

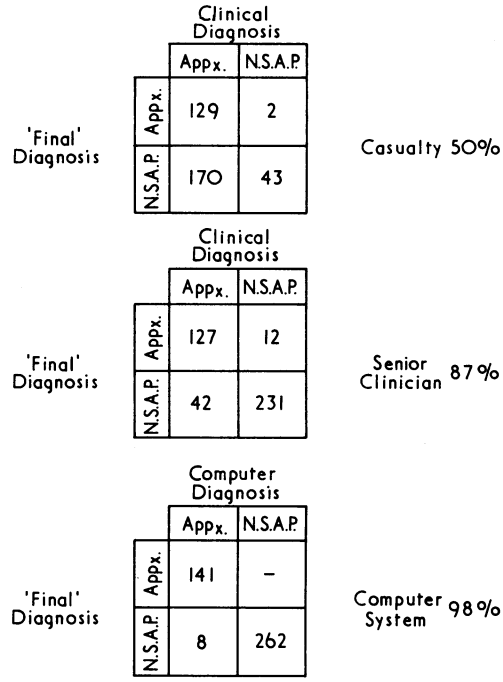

FIG. 2—Comparison of initial clinical and final clinical preoperative diagnostic performance (in discriminating between appendicitis and N.S.A.P.) with performance of computer-aided system. (This simplified diagram ignores other diagnoses made by clinicians or computer-aided system).

PERFORMANCE OF CLINICIANS

We remarked in our earlier report that the performance of clinicians was best reflected in the effectiveness of the decisions which they made—a feature which in turn reflected not only their diagnostic accuracy but also the certainty with which they made their diagnoses. It seems from the data set out in table II that the diagnostic accuracy of registrars and senior clinicians has improved, and though the improvement is not statistically significant it is perhaps of more than passing interest that the diagnostic error rate at registrar level fell by a quarter during the period of the trial.

But this only partially represents a clinician's effectiveness in terms of the decisions he makes. How can these decisions best be assessed? In respect of the acute abdomen there are two ways in which this assessment can be carried out. Firstly, one can estimate the proportion of patients who come to operation for appendicitis *after* their appendix has perforated or an abscess has formed. This is shown in fig. 3. Before the trial in 1969 and 1970 some 40% of appendices had either perforated or formed an abscess by the time the patient reached theatre. During the trial this high proportion gradually fell to around 4% or 5%. The trial was stopped on 1 August 1972, and between then and the end of the year the proportion of appendices which perforated or formed an abscess rose again to 20%.

Secondly one can measure the proportion of patients who undergo "negative laparotomy." Whether one measures this

against the proportion of patients with N.S.A.P. (fig. 4) or as a proportion of the overall laparotomy rate (a more popular but in our view a misleading estimate) it is quite clear that the improvement shown in fig. 3 has not been "bought" at the expense of a higher negative laparotomy rate. Indeed, for what it is worth this seems to have fallen too, from around 25% before the trial to some 6-7% in the later stages of the trial—and there seems also to have been a reversion towards the previous figure once the trial was over.

Discussion

From these data it is difficult to avoid the conclusion that the clinicians' diagnostic and decision-making performance became markedly more effective during the period of the diagnostic trial and reverted towards its previous level afterwards. But at the outset it is necessary to stress that our comparison of the two study periods of the trial does not constitute a formal controlled trial, and this in turn means that there are several theoretical objections which must be raised before any conclusions concerning the clinicians' performances can be accepted. Firstly, the clinicians involved may have changed and the advent of new and "better" (in the sense of more effective) clinicians may have led to the trends shown in figs. 3 and 4. Actually this is not the case; the clinician population remained remarkably stable during the trial, and where a new clinician joined the unit his initial performance went *against* the general improvement. Moreover the population did not change in the crucial post-trial period.

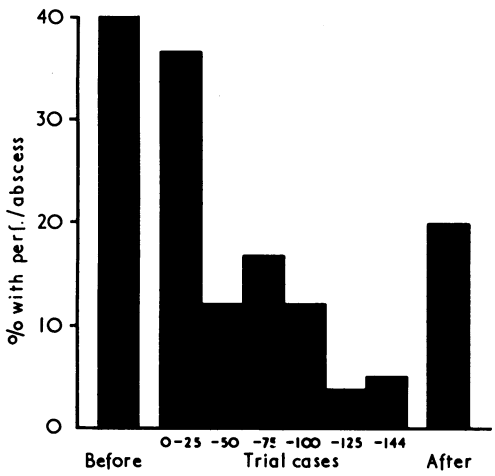

FIG. 3—One method of analysis of clinician's diagnostic effectiveness as regards patients with appendicitis. (For detailed comments see text).

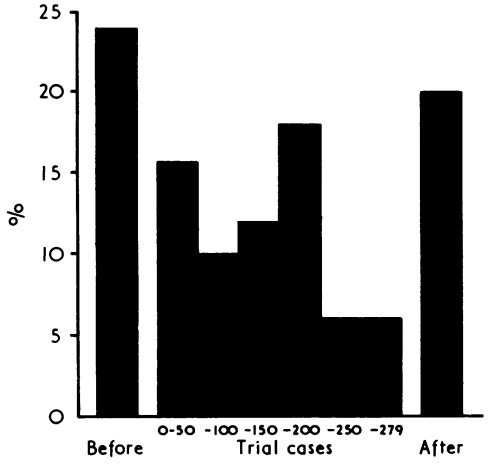

FIG. 4—Proportion of patients with N.S.A.P. coming to negative laparotomy during a three-year period (1969-72).

Perhaps, then, we are merely showing a "learning" process—a group of clinicians gathering experience and improving thereby their decision-making. This, however, is far too facile an explanation since it fails totally to explain the rise in perforation, abscess, and negative laparotomy rates after the trial ceased. Another theoretical possibility is that the improvement seen is due to the fact that the cases became "easier." This too is unlikely. We have studied the patterns of referral closely for such evidence and failed to find it. Thus for example the delay between the onset of symptoms of appendicitis and admission to hospital has not altered over the period of our trial. Also it is worth remembering that the diagnostic accuracy of admitting staff and house surgeons actually fell while the effectiveness of more senior clinicians was improving.

It thus seems not unreasonable to assert that what we have observed does represent a real improvement in the effectiveness of performance of our clinical colleagues, which cannot be accounted for on the basis of staff rotation, "natural" acquisition of experience, or "easier" cases. What then is the reason for this improvement? It would perhaps be gratifying to assert that the computer played some part in this, but if it did it could only be in an indirect way. (One could postulate, for example, that the computer represented an "opponent" to be beaten in a diagnostic "contest"—but this implies that the clinician is so apathetic normally that he requires some such spur to function effectively, and this is an assertion which we do not make.)

Moreover, in this context it is worth pointing out that the computer's predictions for each case were not made available to the clinician until after his primary decision had been made. Equally, it would be idle to say that the clinical information science team had anything directly to do with the improvement—since we studiously refrained from comment until after the clinical team had indicated their primary management of each patient.

But equally it would be idle to ignore the possibility that the discipline forced on our clinical colleagues, the constraint of collecting clinical information to complete a form containing a series of rigidly defined patient attributes, the constant emphasis on the reliability of clinical data collected, and the rapid "feed-back" from the computer-aided system may have combined to account for at least part of the improvement seen during our survey. Sadly, it seems that the effect was short-lived, in that once the trial was over the clinicians' performances reverted to something like their previous level; but it does give some hope that closer attention to clinical information may have something to offer in terms of practical clinical performance.

Finally, it is worth pointing out that in response to the data shown in table II we introduced for a trial period part of the system into the receiving room environment. During December 1972 when this was in operation some 36% of patients with abdominal pain were sent home (as opposed to a corresponding figure of 21% the year before), and the accuracy of diagnosis in cases admitted to hospital rose from around 40% in 1971 to 62% in December 1972. This improvement in performance parallels that seen among our surgical colleagues, and also gives modest encouragement for the future.

We are grateful to Professor J. C. Goligher for his advice and encouragement throughout this study, and to Professor Goligher and Mr. D. Johnston for permission to study patients admitted under their care. We are also grateful to Professor K. Smith and Professor M. Wells, from the computer centre of the University

of Leeds, and to members of their staffs—notably Dr. D. Holdsworth—for their advice and help. Three of us (D.J.L., A.P.M., and J.C.H.) were aided by a grant from the Medical Research Council which we also acknowledge with gratitude. Finally—and particularly—we thank our clinical colleagues for their tolerance and readiness to accept what initially must have seemed a highly esoteric system.

References

[1] Horrocks, Jane C., McCann, A. P., Staniland, J. R., Leaper, D. J., and de Dombal, F. T., *British Medical Journal*, 1972, **2**, 5.
[2] de Dombal, F. T., Leaper, D. J., Staniland, J. R., McCann, A. P., and Horrocks, Jane C., *British Medical Journal*, 1972, **2**, 9.
[3] Shepherd, J., *British Medical Journal*, 1972, **2**, 347.

Occasional Survey

Acute Promyelocytic Leukaemia

J. M. GOLDMAN

British Medical Journal, 1974, 1, 380–382

Summary

Acute promyelocytic leukaemia (A.P.L.) is a rare but important type of acute myeloid leukaemia characterized by major bleeding in association with thrombocytopenia, a specific peripheral blood and bone marrow picture, low plasma fibrinogen, and the presence in the serum of fibrin degradation products. These last abnormalities are related to the disseminated intravascular consumption of coagulation factors with secondary fibrinolysis. A.P.L. requires early recognition and urgent treatment. With optimal management up to half of the patients may achieve complete remission of two years or more. Undoubtedly patients with A.P.L. do especially well when treated in special centres and some patients with A.P.L. now die before the nature of their disease is recognized. Increased familiarity with the problem, which has been known for nearly 20 years, should yield great dividends for those few patients who have this disease.

Abundant evidence indicates that erythrocytes, granulocytes, platelets, and probably monocytes have a common stem cell in the adult bone marrow and the term "acute myeloid leukaemia" (A.M.L.) is often used to embrace all types of acute leukaemia thought to arise from this cell—myeloblastic, myelomonocytic, monocytic, erythroleukaemic, and promyelocytic. The term "acute promyelocytic leukaemia" (A.P.L.) describes a special type of A.M.L. characterized by severe bleeding manifestations which are frequently fatal, the presence in the peripheral blood and bone marrow of atypical promyelocytes containing large azurophilic granules, thrombocytopenia, and abnormalities in clotting test results that reflect disseminated intravascular coagulation. Recently the treatment of A.P.L. has improved greatly and the incidence of complete remission now approaches or equals that seen in other types of A.M.L. Generally these remissions appear to be much longer than those in other types of A.M.L. and are frequently measured in years. The analogy with acute lymphoblastic leukaemia, in which cure appears to depend on adequate chemotherapy during remission and to be closely related to the total duration of disease-free survival suggests that it may prove easier to eradicate residual disease in patients with A.P.L. in remission than in other types of A.M.L.

The first description of acute leukaemia associated with hypofibrinogenaemia is attributed to Croizat and Favre-Gilly.[1] Eight years later Hillestad[2] described a special type of A.M.L. characterized by the presence of atypical promyelocytes, hypofibrinogenaemia, and a severe bleeding tendency and introduced the term "acute promyelocytic leukaemia." Later the low fibrinogen was found to be due to rapid destruction and the idea that the clinical picture was due to fibrinolysis or to increased intravascular consumption was formed.[3-4]

Many cases have now been diagnosed and A.P.L. probably represents about 6% of all cases of acute leukaemia or 13% of all cases of A.M.L.[5 8-7] This review summarizes the clinical, haematological, and coagulation features of A.P.L. and the important improvements in treatment and prognosis that have recently occurred.

Clinical Features

There are no clinical criteria for a positive diagnosis of A.P.L., yet it differs from other varieties of A.M.L. in the frequency of bleeding manifestations and the relative absence of enlarged viscera. Thus, a high proportion of patients present with purpura, epistaxis, haematuria, or bleeding from gums or other gastrointestinal sites. Death often results from cerebral haemorrhage. The liver and spleen are less frequently palpable in A.P.L. than in other types of A.M.L. and lymphadenopathy is usually absent. Gum hypertrophy and infiltration of the skin are unusual in A.P.L.

Haematological Features

The total peripheral blood leucocyte count is variable but low counts are much commoner at the time of presentation of A.P.L. than in other types of A.M.L. This may reflect relatively earlier diagnosis and the leucocyte count may rise rapidly after diagnosis if treatment is not started. Blast cells are usually present in the peripheral blood but the predominant cell is frequently an atypical promyelocyte. This cell differs from a normal promyelocyte in that nucleoli are often inconspicuous; the nucleus is more likely to be indented, folded, or bilobed and is frequently double; while the cytoplasm contains many large densely-staining azurophilic granules which may overlie and occlude the

Hammersmith Hospital, London W12 0HS
JOHN M. GOLDMAN, B.M., M.R.C.P., Honorary Consultant Physician